

Forecasting Stock Market using Multiple Linear Regression

Yee Mon Khaing, Myint Myint Yee, Ei Ei Aung

University of Computer Studies, Maubin, Myanmar

How to cite this paper: Yee Mon Khaing | Myint Myint Yee | Ei Ei Aung "Forecasting Stock Market using Multiple Linear Regression"

Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-5, August 2019, pp.2174-2176, <https://doi.org/10.31142/ijtsrd27819>



IJTSRD27819

Copyright © 2019 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



1.1. SPSS

The "Statistical Package for the Social Sciences" (SPSS) is a package programs for manipulating, analyzing, and presenting data. SPSS is widely used by market researchers, health researchers, survey companies, government entities, education researchers, marketing organizations, data miners, and many more for the processing and analyzing of survey data. [2]

1.2. Regression

Regression is a powerful statistical method used in education, finance, investing and other disciplines that allow estimating the relationships between one dependent variable (usually denoted by Y) and one or more independent. As with most statistical analyses, the goal of regression is to summarize observed data as simply, usefully, and elegantly as possible. The two basic types of regression are simple *linear regression* and *multiple linear regression*, although there are *non-linear regression* methods for more complicated data and analysis. Simple linear regression uses one independent variable to predict the outcome of the dependent variable whereas multiple linear regression uses two or more independent variables to predict the outcome of the dependent variable. [4]

2. Methodology [3][5][6]

2.1. Multiple Linear Regression

Multiple linear regression (MLR), known as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. Multiple regression is a powerful technique used for predicting the unknown value of a variable from the

ABSTRACT

Regression is one of the most powerful statistical methods used in business and marketing researches. This paper shows the important instance of regression methodology called Multiple Linear Regression (MLR) and proposes a framework of the forecasting of the Stock Index Price, based on the Interest Rate and the Unemployment Rate. This paper was applied the aid of the Statistical Package for Social Sciences (SPSS) version 23 and PYTHON version 3.7.

KEYWORDS: MLR, SPSS, Stock Index Price, Interest Rate, Unemployment Rate

1. INTRODUCTION

Knowledge of stock exchange behavior is very important for investors, traders and financial manager. The investors, traders and investment firms need to predict the stock index prices and want to know the time to buy and sell. But, it isn't simply way to predict the behavior of the stock market. Stock index price is influenced by a number of factors in stock market. A number of factors can cause rises and falls in share prices through gradual changes or sharp spikes. Some factors can play a big effect on individual stocks. The main purpose of this paper is to find out that the interest rates and unemployment rates, those factors can or can't influence the stock index prices. This paper uses multiple linear regression to predict the stock index price (the dependent variable) of a fictitious economy by using two independent/input variables: the interest rate and employment rate. It applies multiple linear regression method for data forecasting and ANOVA for data significant.

known value of two or more variables- also called the predictors.

$$y = mx_1 + mx_2 + mx_3 + b$$

where,

y = the dependent variable of the regression equation

m = slope of the regression equation

x₁ = first independent variable of the equation

x₂ = second independent variable of the equation

x₃ = third independent variable of the equation

b = constant of the equation

2.2. R Square

R-Square is a statistical measurement in a regression that calculates the proportion of variance in a dependent variable that is explained by an independent variable or variables. R-square tells how well the data fit the regression model (the goodness of fit). R-squared can take any values between 0 and 1. R-square is better if the values are closer to 1.

R Square Formula = r^2

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

where,

r = the correlation coefficient

n = number in the given dataset

x = first variable in the context

y = second variable

2.3. ANOVA Table

ANOVA is the short form of analysis of variance. ANOVA is a statistical tool which is generally used on random variables. It involves group not directly related to each other in order to find whether exist any common means.

2.4. Significance F-Value

F-Test is any test that uses F-distribution. F value is a value on the F distribution. Various statistical tests generate an F value. The value can be used to determine whether the test is statistically significant. In order to compare two variances, one has to calculate the ratio of the two variances:

$$F = \sigma_1^2 / \sigma_2^2$$

where,

σ_1^2 = larger sample variance

σ_2^2 = smaller sample variance

2.5. P-Value

P is a statistical measure that helps researchers to determine whether their hypothesis is correct. It helps determine the significance of result. P-Value is a number between 0 and 1.

Calculating P-Value from a Z Statistic statistic z

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

where,

\hat{p} is Sample Proportion

p_0 is assumed Population Proportion in the Null Hypothesis

n is the Sample Size

3. Testing

Table-1: Sample Data

| Year | Month | InterestRate | UnemploymentRate | StockIndexPrice |
|------|-------|--------------|------------------|-----------------|
| 2017 | 12 | 2.75 | 5.3 | 1464 |
| 2017 | 11 | 2.5 | 5.3 | 1394 |
| 2017 | 10 | 2.5 | 5.3 | 1357 |
| 2017 | 9 | 2.5 | 5.3 | 1293 |
| 2017 | 8 | 2.5 | 5.4 | 1256 |
| 2017 | 7 | 2.5 | 5.6 | 1254 |
| 2017 | 6 | 2.5 | 5.5 | 1234 |
| 2017 | 5 | 2.25 | 5.5 | 1195 |
| 2017 | 4 | 2.25 | 5.5 | 1159 |
| 2017 | 3 | 2.25 | 5.6 | 1167 |
| 2017 | 2 | 2 | 5.7 | 1130 |
| 2017 | 1 | 2 | 5.9 | 1075 |
| 2016 | 12 | 2 | 6 | 1047 |
| 2016 | 11 | 1.75 | 5.9 | 965 |
| 2016 | 10 | 1.75 | 5.8 | 943 |
| 2016 | 9 | 1.75 | 6.1 | 958 |
| 2016 | 8 | 1.75 | 6.2 | 971 |
| 2016 | 7 | 1.75 | 6.1 | 949 |
| 2016 | 6 | 1.75 | 6.1 | 884 |
| 2016 | 5 | 1.75 | 6.1 | 866 |
| 2016 | 4 | 1.75 | 5.9 | 876 |
| 2016 | 3 | 1.75 | 6.2 | 822 |
| 2016 | 2 | 1.75 | 6.2 | 704 |
| 2016 | 1 | 1.75 | 6.1 | 719 |

The table provides us the data needed to perform the multiple regression analysis. We can predict that there is a relation between Stock Index Price (Output) and Interest Rate and Unemployment Rate (Input).

Table-2: Regression Values
SUMMARY OUTPUT

| Regression Statistics | |
|-----------------------|--------|
| Multiple R | 0.947 |
| R Square | 0.898 |
| Adjusted R Square | 0.888 |
| Standard Error | 70.562 |
| Observations | 24 |

R-square is better if the values are closer to 1. In the table, the result of R Square is 0.898 that's good. Therefore, the proportion of the variance is 90% for Stock Index Price that is explained by Interest Rate and Unemployment Rate.

Table-3 and Table-4: ANOVA Table

| ANOVA | | | | |
|------------|----|------------|-----------|--------|
| | df | SS | MS | F |
| Regression | 2 | 51587.161 | 45842.580 | 92.073 |
| Residual | 21 | 104558.570 | 4979.024 | |
| Total | 23 | 156145.731 | | |

| | df | SS | MS | F | P-value | t-Stat | Standard Error | Lower 95% | Upper 95% |
|------------------|----|----------|----------|-------|---------|--------|----------------|-----------|-----------|
| Intercept | 1 | 4798.404 | 4798.404 | 2020 | 0.000 | 20.20 | 899.143 | -71.335 | 3633.493 |
| InterestRate | 1 | 345.540 | 345.540 | 3.103 | 0.005 | 3.103 | 111.367 | -113.340 | 577.140 |
| UnemploymentRate | 1 | 250.147 | 250.147 | 2.121 | 0.046 | -2.121 | 117.950 | -495.337 | 435.037 |

As Table 3
Significance F and P-values

This table tests the statistical significance of the independent variables as predictors of the dependent variable. The last column of the table shows the results of an overall F test. The F statistic (92.073) is big, and the p value (0.000) is small. This indicates that one or both independent variables have explanatory power beyond what would be expected by chance.

As table-4, Significance of Regression Coefficients

The coefficients table shows the following information each coefficient: its value, its standard error, a t-statistic, and the significance of the t-statistic. In this table, the t-statistics for the interest rate and the unemployment rate are both statistically significant at the 0.05 level. This means that the interest rate contributes significantly to the regression after effects of unemployment rate are taken into account. And the unemployment rate contributes significantly to the regression after effects of the interest rate are taken into account.

Table-5: RESIDUAL OUTPUT

| Observation | Predicted StockIndexPrice | Residuals |
|-------------|---------------------------|-----------|
| 1 | 1422.862 | 41.138 |
| 2 | 1336.477 | 57.523 |
| 3 | 1336.477 | 20.523 |
| 4 | 1336.477 | -43.477 |
| 5 | 1311.463 | -55.463 |
| 6 | 1261.433 | -7.433 |
| 7 | 1286.448 | -52.448 |
| 8 | 1200.063 | -5.063 |
| 9 | 1200.063 | -41.063 |
| 10 | 1175.048 | -8.048 |
| 11 | 1063.649 | 66.351 |
| 12 | 1013.619 | 61.381 |
| 13 | 988.605 | 58.395 |
| 14 | 927.234 | 37.766 |
| 15 | 952.249 | -9.249 |
| 16 | 877.205 | 80.795 |
| 17 | 852.190 | 118.810 |
| 18 | 877.205 | 71.795 |
| 19 | 877.205 | 6.795 |
| 20 | 877.205 | -11.205 |
| 21 | 927.234 | -51.234 |
| 22 | 852.190 | -30.190 |
| 23 | 852.190 | -148.190 |
| 24 | 877.205 | -158.205 |

The result of coefficients can use to do a forecast. The regression line is: $y = \text{Stock Index Price} = 1798.404 + 345.540 * \text{Interest Rate} - 250.147 * \text{Unemployment Rate}$. In other words, for each unit increase in Interest Rate, Stock Index Price increase with 345.540 units. For each unit increase in Unemployment Rate, Stock Index Price decreases with 250.147 units. This is valuable information.

*Example 1, if Interest Rate equals 1 and Unemployment Rate equals 5, you might be able to achieve the Stock Index Price of $1798.404 + 345.540 * 1 - 250.147 * 5.5 = 768$.*

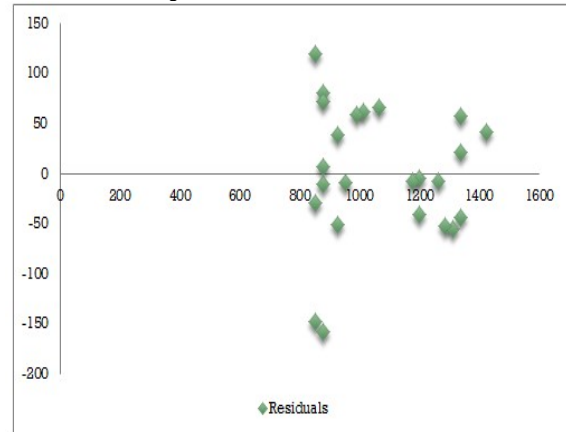
*Example 2, if Interest Rate equals 3.5 and Unemployment Rate equals 5, you might be able to achieve the Stock Index Price of $1798.404 + 345.540 * 3 - 250.147 * 5.5 = 1459$.*

As Example 1 and 2, when interest rate goes up, the stock index price also goes up (here we still have a linear relationship with a positive slope).

*Example 3, if Interest Rate equals 1.5 and Unemployment Rate equals 5, you might be able to achieve the Stock Index Price of $1798.404 + 345.540 * 3 - 250.147 * 5 = 1584$.*

*Example 4, if Interest Rate equals 1.5 and Unemployment Rate equals 6.5, you might be able to achieve the Stock Index Price of $1798.404 + 345.540 * 3 - 250.147 * 6.5 = 1209$.*

As Example 3 and 4, when unemployment rate goes up, the stock index price goes down (here we still have a linear relationship, but with a negative slope).

Graph-1: Residuals Values

As table-5 and graph-1 Residuals

The residuals show you how far away the actual data points are from the predicted data points (using the equation).

*For example, the first data point equals 1464. Using the equation, the predicted data point equals $1798.404 + 345.540 * 2.75 - 250.147 * 5.3 = 1422.862$, giving a residual of $1464 - 1422.862 = 41.138$.*

*For another example, the second data point equals 1394. Using the equation, the predicted data point equals $1798.404 + 345.540 * 2.5 - 250.147 * 5.3 = 1336.477$, giving a residual of $1394 - 1336.477 = 57.523$.*

4. Conclusion

Regression is one of the most powerful statistical methods used in business and marketing researches and an important instance of regression methodology is Multiple Linear Regression (MLR). SPSS data analysis tools are valuable in education, business and marketing fields. It is very good for presentation report by graphical design. This main purpose of this analysis is to know to what extent is the Stock Index Price influenced by the two independent variables, the Interest Rate and the Unemployment Rate. This paper intends to support the traders, investors and financial managers to predict the behavior of the stock market.

References

- [1] IBM SPSS Statistics 24 Algorithms pdf book [book style]
- [2] A handbook of statistical analyses using SPSS / Sabine, Landau, Brian S. Everitt, ISBN 1-58488-369-3 [book style]
- [3] <https://www.exceleasy.com>
- [4] <https://www.investopedia.com/terms/r/regression.asp>
- [5] <https://www.wallstreetmojo.com>
- [6] <https://datatofish.com/multiple-linear-regression-python/>